

LISTING OF CLAIMS

1. (Previously Presented) A computer medium of sound or image recognition comprising:

one or more sensors or receivers responsive to signals;

a computer operatively coupled to the one or more sensors, the computer comprising a central processing unit;

one or more memories, at least one of the one or more memories storing a software program comprising the steps of:

defining a plurality of distributions of known database records onto respective training and testing subsets;

training and testing a first generation set of prediction algorithms using the plurality of distributions of the database records, each of said prediction algorithms being associated with a first different distribution of said database records;

assigning a fitness score to each of the prediction algorithms;

feeding the set of prediction algorithms to an evolutionary algorithm which generates a set of one or more second generation prediction algorithms and assigns a fitness score to each;

continuing to feed each generational set of prediction algorithms to the evolutionary algorithm until a termination event occurs, wherein said termination event is at least one of:

a prediction algorithm generated with a fitness score equal to or exceeding a defined minimum value,

the maximum fitness score of successive generational sets of prediction algorithms converging to a given value, or

a certain number of generations having been generated;

selecting a prediction algorithm having a best fitness score; and

using the distribution of database records associated with said selected prediction algorithm in performing supervised learning, said supervised learning

including training and testing of prediction algorithms to obtain a trained prediction algorithm;

generating a population of prediction algorithms, wherein each of said prediction algorithms is trained and tested according to a second different distribution of the records of the data set in the complete database onto a training data set and a testing data set,

each second different distribution being created as one of a random or pseudorandom distribution,

each prediction algorithm of said population being trained according to its own distribution of records of the training set and being validated in a blind way according to its own distribution on the testing set, and

a score reached by each prediction algorithm being calculated in the testing phase representing its fitness;

providing an evolutionary algorithm which combines the different models of distribution of the records of the complete data set in a training and in a testing set, which sets are represented each one by a corresponding prediction algorithm trained and tested on the basis of said training and testing data set according to the fitness score calculated in the previous step for the corresponding prediction algorithm,

the fitness score of each prediction algorithm corresponding to one of the different distributions of the complete data set on the training and the testing data sets being the probability of evolution of each prediction algorithm or of each said distribution of the complete data set on the training and testing data sets;

repeating the evolution of the prediction algorithm generation for a finite number of generations or till the output of the genetic algorithm converges to a best solution and/or till the fitness value of at least some prediction algorithm related to an associated data records distribution has reached a desired value; and

setting the data records distribution for the best solution as the optimized training and testing subsets for training and testing prediction algorithm; and

an output system providing an indication of the signals detected by the one or more sensors.

2. (Canceled)

3. (Previously Presented) The computer medium according to claim 1, wherein the software program further comprises the step of associating a distribution variable to each record of the data set, which is binary and which has at least two statuses, one of the two statuses being associated with the inclusion of the record in the training set and the other one of the two statuses in the testing set.

4. (Previously Presented) The computer medium according to claim 1, wherein the prediction algorithm is an artificial neural network.

5. (Previously Presented) The computer medium according to claim 1, wherein the prediction algorithm is a classification algorithm.

6. (Previously Presented) The computer medium according to claim 1, wherein once an optimum distribution has been computed, the optimized training data subset is made equal to a complete data set, the individuals included in the training subset being distributed onto a new training set and onto a new testing set each having about half of the records of the original optimized training set, while the originally optimized testing set is used as a third data subset for validation purposes.

7. (Previously Presented) The computer medium according to claim 6, wherein the distribution of the data of the originally optimized training set onto the new training and new testing set is optimized through a pre-processing phase including the steps of said method for optimizing a database of sample records, said records being records in the originally optimized training set.

8. (Previously Presented) The computer medium according to claim 1, wherein different choices of the structure of the training subset and the structure of the testing subset comprise different selections of the number of input variables of the data records of the

database, which selections include leaving out at least one variable from the entire input variable set forming each record, the records of the database comprising a certain number of known input variables and a certain number of known output variables.

9. (Previously Presented) The computer medium according to claim 8, further comprising the following steps:

defining a distribution of data from the complete data set onto a training data set and onto a testing data set;

generating a population of different prediction algorithms each one having a training and/or testing data set in which only some variables have been considered among all the original variables provided in the data sets, each one of the prediction algorithms being generated through a different selection of variables;

carrying out learning and testing of each prediction algorithm of the population and evaluating the fitness score of each prediction algorithm;

applying an evolutionary algorithm to the population of prediction algorithms for achieving new generations of prediction algorithms;

for each generation of new prediction algorithms, representing a new different selection of input variables, testing or validating the best prediction algorithm according to the best hypothesis of input variables selection; and

evaluating a fitness score and promoting the prediction algorithms, representing the selections of input variables which have the best testing performances and the minimum input variables, for the processing of the new generations.

10. (Previously Presented) The computer medium according to claim 8, further comprising a preprocessing phase, including the steps of said method for optimizing a database of sample records, for selecting the most predictive input variables.

11. (Previously Presented) The computer medium according to claim 1,
in which different choices of the structure of the training subset and the structure of the testing subset comprise different selections of the number of input variables of the data records

of the database, which selections include leaving out at least one, variable from the entire input variable set forming each record, the records of the database comprising a certain number of known input variables and a certain number of known output variables,

and further comprising a pre-processing phase, including the steps of said method for optimizing a database of sample records, for selecting the most predictive input variables,

wherein the database subjected to the pre-processing phase of input variable selection is a training subset and a testing subset processed with said method.

12. (Previously Presented) The computer medium according to claim 1, wherein the complete database the distribution of the records of which has to be optimized has data records having a selected number of input variables, the selection being carried out with said method, and wherein different choices of the structure of the training subset and the structure of the testing subset comprise different selections of the number of input variables of the data records of the database, which selections consist in leaving out at least one variable from the entire input variable set forming each record, the records of the database comprising a certain number of known input variables and a certain number of known output variables.

13. (Previously Presented) The computer medium according to claim 1, wherein a pre-processing phase for optimizing the distribution of the records on a training subset and a testing subset and for selecting the most predictive input variables, is carried out alternatively one to the other several times.

14. (Previously Presented) The computer medium according to claim 1, wherein the evolutionary algorithm is a genetic algorithm with the following evolutionary rules:

an average health value of the population is computed as a function of the fitness values of each single individual in the population;

coupling, recombination of genes and mutation of genes are carried out in a differentiated manner depending on a comparison between the fitness of each individual of the couple and the average health value of the entire population to which the individuals belong;

individuals having a fitness value lower or equal to the average health of the entire

population are not excluded from the creation of new generations but are marked out and entered in a vulnerability list; and

the number of subjects entered in the vulnerability list defines the number of possible marriages.

15. (Previously Presented) The computer medium according to claim 14, wherein for coupling purposes and for generation of children at least one parent individuals must have a fitness value greater than the average health value of the population.

16. (Previously Presented) The computer medium according to claim 14, wherein each couple of individuals are adapted to generate offsprings having a fitness different from the average health if the fitness of one them at least is greater than the average fitness, the offsprings of each marriage occupying the places of subjects entered in the vulnerability list and marked out, so that a weak individual can continue to exist through his own children.

17. (Previously Presented) The computer medium according to claim 14, wherein coupling between individuals having a very low fitness value and a very high fitness value are not allowed.

18. (Previously Presented) The computer medium according to claim 14, wherein the following recombination rules of the genes of the coupled parent individuals are considered in the case the parents individuals have no common genes:

the health of father and mother individuals are greater than the average health of the entire population;

the crossover is a classical crossover according to which the genes of the father and of the mother individuals are substituted one with the other starting from a certain crossover point;

the health of father and mother individuals are lower than the average health of the entire population, in this case the two children are formed through rejection of the parents' genes they will receive by the crossover process;

the health of one of the parents is less than the average health of the entire population

while the health of the other parent is greater than the average health of the entire population, in this case only the parents whose health is greater than the average health of the entire population will transmit their genes, while the genes of the parent having an health lower than the average health of the entire population are rejected.

19. (Previously Presented) The computer medium according to claim 18, wherein each gene is characterized by a status level, and wherein gene rejection comprises modifying the status of the genes from one status level to a different status level.

20. (Previously Presented) The computer medium according to claim 18, wherein a modified crossover of the genes of the parent individuals is carried out when the parent individuals have part of the genes that coincide, this modified crossover providing for generating an offspring in which the genes selected for crossover are the most effective ones of the parents.

21. (Previously Presented) The computer medium according to claim 14, wherein the individuals are the different prediction algorithms representing a corresponding different initial random distribution of data records onto the testing and the training data set, and wherein the genes consist in the binary status variable of association of each record to the training and to the testing subset.

22. (Previously Presented) The computer medium according to claim 14, wherein the individuals are the prediction algorithms each one representing a different training and testing data set, the difference residing in a different selection of input variables for each different training and testing subset, and wherein the genes comprise a different selection variable which is provided for each input variable in the different training and testing subsets, the selection variable being a parameter indicating the presence/absence of each corresponding input variable in the records of each data set.

23. - 26. (Canceled)

27. (Previously Presented) The computer medium according to claim 1, wherein the output is an indication of a shape of an object generating or reflecting electromagnetic waves, and/or the distance and/or the identity of the object.

28. (Previously Presented) The computer medium according to claim 1, wherein the known database records comprise acoustic signals emitted by one or more objects or one or more living beings making part of a typical environment in which the method is performed or data relating to one or more images of one or more objects or one or more living beings that are part of the typical environment, and/or identity and/or meaning of objects to which the said acoustic signals or image data are related and/or from which said acoustic signals or image data are generated.

29. (Previously Presented) The computer medium according to claim 27, wherein the computer medium is a specialized system for image pattern recognition having artificial intelligence utilities for analyzing an image in the form of a array of image data records, each image data record being related to a zone or point or unitary area or volume of a two or three dimensional visual image, the visual image being formed by an array of pixels or voxels and utilities for indicating for each image data record a certain quality among a plurality of known qualities of the image data records;

wherein the one or more sensors or receivers receive arrays of digital image data records or generate an array of digital image data records from an existing image;

wherein at least one of one of the one or more memories stores said digital image data array, and

wherein the output system indicates for each image data record of the image data array a certain quality chosen by the processing unit in carrying out the image pattern recognition algorithm in the form of the said software program.

30. (Canceled)

31. (Previously Presented) A computer medium according to claim 38, wherein an optimization of the distribution of the records of the original database in a training dataset and in a testing dataset is carried out in one of a pre processing and a post processing phase.

32. – 35. (Canceled)

36. (Previously Presented) The computer medium according to claim 1, wherein the signals are electromagnetic waves in the acoustic or visible range.

37. (Previously Presented) The computer medium according to claim 1, wherein the software program further comprises a preprocessing phase comprising the steps of:

defining a plurality of distributions of the records of the optimized training subset onto new training and testing subsets;

training and testing a new generation set of prediction algorithms using the new training and testing subsets;

assigning a fitness score to each prediction algorithm in the new generation of prediction algorithms;

defining a new optimized training subset and a new optimized testing subset;

identifying a new optimized training subset and a new optimized testing subset as the training and testing subsets corresponding to the prediction algorithm having the highest fitness score; and

employing the optimized testing subset as a validation set.

38. (Previously Presented) A computer medium for producing a microarray for genotyping, the computer medium comprising:

a computer comprising a central processing unit;

one or more memories, at least one of the one or more memories storing a database of experimentally determined data in which each record relates to a known clinical or experimental case of a sample population of cases, the data comprising a number of input variables

corresponding to the presence/absence of a predetermined number of polymorphisms and/or mutations and/or equivalent genes of a number of theoretically probable relevant genes, said certain predetermined number of polymorphisms and/or genes forming a set, and the data further comprising one or more related output variables corresponding to the certain biological or pathologic condition of the clinical and experimental cases of the sample population;

at least one of the one or more memories storing a software program defining a number of theoretically relevant genes or alleles or polymorphisms relevant for a biologic condition, the software program comprising the steps of:

determining a selection of a subset of the set of certain predetermined number of polymorphisms and/or genes by testing the association of the genes or polymorphisms and the biological or pathological condition by mathematical tools comprising a prediction algorithm applied to the database;

defining a plurality of distributions of the database onto respective training and testing subsets;

training and testing a first generation set of prediction algorithms using the plurality of distributions of the database, each of said prediction algorithms being associated with a first different distribution of records of the database;

assigning a fitness score to each of the prediction algorithms;

feeding the set of prediction algorithms to an evolutionary algorithm which generates a set of one or more second generation prediction algorithms and assigns a fitness score to each;

continuing to feed each generational set of prediction algorithms to the evolutionary algorithm until a termination event occurs, wherein said termination event is at least one of:

a prediction algorithm generated with a fitness score equal to or exceeding a defined minimum value,

the maximum fitness score of successive generational sets of prediction algorithms converging to a given value, or

a certain number of generations having been generated;

selecting a prediction algorithm having a best fitness score; and

using the distribution of database associated with said selected prediction algorithm in performing supervised learning, said supervised learning including training and testing of prediction algorithms to obtain a trained prediction algorithm;

generating a population of prediction algorithms, wherein each of said prediction algorithms is trained and tested according to a second different distribution of the records of the data set in the complete database onto a training data set and a testing data set,

each second different distribution being created as one of a random or pseudorandom distribution,

each prediction algorithm of said population being trained according to its own distribution of records of the training set and being validated in a blind way according to its own distribution on the testing set, and

a score reached by each prediction algorithm being calculated in the testing phase representing its fitness;

providing an evolutionary algorithm which combines the different models of distribution of the records of the complete data set in a training and in a testing set, which sets are represented each one by a corresponding prediction algorithm trained and tested on the basis of said training and testing data set according to the fitness score calculated in the previous step for the corresponding prediction algorithm,

the fitness score of each prediction algorithm corresponding to one of the different distributions of the complete data set on the training and the testing data sets being the probability of evolution of each prediction algorithm or of each said distribution of the complete data set on the training and testing data sets;

repeating the evolution of the prediction algorithm generation for a finite number of generations or till the output of the genetic algorithm converges to a best solution and/or till the fitness value of at least some prediction algorithm related to an associated data records distribution has reached a desired value; and

setting the data records distribution for the best solution as the optimized training and testing subsets for training and testing prediction algorithm; and
an output system responsive to the received information.